Current Biology, Vol. 13, R86–R87, February 4, 2003, ©2003 Elsevier Science Ltd. All rights reserved. PII S0960-9822(03)00032-0

Human Prehistory: The Message from Linkage Disequilibrium

Dispatch

John F.Y. Brookfield

The vast amount of information being gathered about human DNA sequence variation raises the question of what these data can tell us about events in our past. A new way has been found by which patterns of linkage disequilibrium can be used to detect the effects of natural selection in human prehistory.

Given the evolutionary timescale, changes in allele frequency that are driven by selection are unlikely to be directly observable in one or a few human lifetimes. One success of the ecological genetics tradition was that for some polymorphisms, such as the dark and light morphs of the peppered moth *Biston betularia*, frequency changes were sufficiently rapid for us to monitor. The explanation for this unexpected good fortune was that, as with the spread of insecticide resistance alleles in other insects, the peppered moth population had been subject to a profound and man-made environmental change that had reversed selective differentials. For humans, however, with their very long generation times, allele frequencies change slowly. If we wish to detect selectively driven allele frequency changes in human populations, all that we have is the snapshot of the variation present in contemporaneous human populations. But a new study by Sabeti *et al.* [1] has now revealed a new way of using patterns of linkage disequilibrium to detect the effects of natural selection in human prehistory.

In interpreting patterns of sequence variation, observations are typically compared to the predictions of the 'standard neutral model' [2]. In this model, it is assumed that there is a single randomly mating (panmictic) population of constant effective population size; that all mutations are neutral; and that mutations always occur in different bases in the DNA (the so-called 'infinite sites' assumption). Such a model will yield predictions of the level of genetic diversity in the population and the frequency distribution of variable sites. And if the rate of recombination is known, the model predicts the expected degree of linkage disequilibrium.

Linkage disequilibrium is a measure of the association in a population between polymorphic sites in the DNA. A new mutation will arise in the background of a particular haplotype, and so in a state of linkage disequilibrium; but as the mutation spreads, its linkage disequilibium with other sites in the haplotype diminishes as it is separated from them by recombination. In the standard neutral model at equilibrium, linkage disequilibrium decays with map distance along the chromosome at a rate determined by the effective population size. For example, the decay of linkage disequilibrium with distance along human chromosomes is found to be more rapid in African populations, which have larger effective population sizes than other human groups [3]. Selection can be inferred from patterns of variation that are inconsistent with these neutral predictions.

One type of selected change that may have happened in human populations is that seen in the peppered moth — the rapid spread of a new mutation as a result of changed environmental conditions. In humans, one environmental change likely to have led to the selection of new alleles at various loci was the increase in infectious disease that followed the development of agriculture.

When a new allele arises and is spread by selection, the allele can reach a high frequency in the population so rapidly that recombination does not have a chance to destroy the linkage disequilibrium present when the mutation is created. In their new study, Sabeti *et al.* [1] thus argue that the sign of a recently selectively favoured allele is one that is at a high frequency but still shows high linkage disequilibrium. To test whether this effect can be discerned, they examined the haplotype structure around two genes: *glucose 6-phosphate dehydrogenase* (*G6PD*) and the CD40 ligand gene (*TNFSF5*). At each locus, there was prior evidence that one allele is favoured because it confers resistance to malaria [4,5]. For the allele *G6PD-202A*, this evidence itself included a prior study of linkage disequilibrium using microsatellite markers [4]. Each favoured allele is only found among Africans, and the authors carried out a survey using 60 people from the Beni population, 87 Yoruba and 83 Shona.

The method starts by examining the 'focal' gene whose alleles confer fitness differences. Single nucleotide polymorphisms (SNPs) are used to define a small number of haplotypes for this gene, one of which is associated with the protective allele. Now, further SNPs are identified in the flanking sequences, up to 500 kilobases upstream and downstream of the focal gene. With complete linkage disequilibrium throughout the region, every chromosome with a given haplotype at the focal gene would have an identical genotype for each SNP both upstream and downstream. In reality, as one goes further away from the focal gene, variation starts to appear.

Sabeti *et al.* [1] define a statistic that they call the 'extended haplotype homozygosity' (EHH), which can be measured for each haplotype for any distance from the focal gene. This is the probability that two randomly chosen chromosomes from the sample that share the same focal gene haplotype also show identical haplotypes for their SNP patterns in the surrounding DNA. Clearly, the EHH will decrease with distance from the focal gene, and the rate of decay can be compared between haplotypes. The authors found that, at each gene, a selected haplotype shows

Institute of Genetics, University of Nottingham, Queens Medical Centre, Nottingham, NG7 2UH, UK.

an EHH that decays much more slowly with distance than does the EHH for the other haplotypes. This is a sign of the recent spread of that haplotype.

While the rate of decay of linkage disequilibrium with distance depends on the recombination rate, which is not known, recombination will be the same for all haplotypes, and thus the selectively advantageous haplotype can be distinguished by comparing its pattern of EHH decay with that of the other haplotypes at the locus. The significance of the high EHH was also assessed by simulation based on the standard neutral model, and the probability of it remaining as high as was seen for the selected haplotypes was calculated as being below 0.1% for each of the two selected alleles.

Although at first sight these results merely confirm the selective advantage of these two malaria-resistance alleles, the true importance of the study — and of another recent study that also detected selection through linkage disequilibrium [6] — is that it provides a screening process for selection that could be applied to raw sequence variation data. The idea is that we can use the EHH in the new data sets showing the human haplotype structure [7,8] to scan the genome looking for alleles that have recently arisen but are now at high frequency. The time scale is important. This method could not detect selectively driven increases in allele frequency that occurred much earlier in human prehistory — their flanking sequences would have reached linkage equilibrium by now.

There remains a general problem with tests detecting selection that compare human diversity with predictions of the standard neutral model, which is that the human population has been neither constant in size nor panmictic [2]. It has expanded and also shows population substructure. With substructure, drawing conclusions from allele frequencies is always dangerous. In the work of Sabeti *et al.* [1], neither gene showed variation in haplotype frequencies between the three African populations sampled, but both selected alleles were absent in Americans of European ancestry and in Asians. This is unsurprising. An allele that was generated recently — within the last thousand generations — and has risen to high frequencies will almost certainly be geographically clustered. This makes the sampling strategy particularly important. Evidence for selection comes not from a high EHH *per se* but from an EHH that is high in relation to the haplotype's frequency. Had the selected haplotype frequencies been lower, the EHH values for these haplotypes would not have been high enough to indicate selection. When haplotype frequencies vary between samples, the mean frequencies across samples may not be powerful in detecting the signature of selection.

References

1. Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z.P., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J. *et al*. (2002). Detecting recent positive selection in the human genome from haplotype structure. Nature *419*, 832–837.
2. Przeworski, M., Hudson, R.R. and Di Rienzo, A. (2000). Adjusting the focus on human variation. Trends Genet. *16*, 296–302.
3. Frisse, L., Hudson, R.R., Bartioszewicz, A., Wall, J.D., Donfack, J. and Di Rienzo, A. (2001). Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. Am. J. Hum. Genet. *69*, 831–843.
4. Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J. *et al*. (2001). Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. Science *293*, 455–462.
5. Sabeti, P., Usen, S., Farhadian, S., Jallow, M., Doherty, T., Newport, M., Pinder, M., Ward, R. and Kwiatkowski, D. (2002). CD40L association with protection from severe malaria. Genes Immun. *3*, 286–291.
6. Toomajian, C. and Kreitman, M. (2002). Sequence variation and haplotype structure at the human *HFE* locus. Genetics *161*, 1609–1623.
7. Riech, D., Cargill, M., Bolk, S., Ireland, J., Sabeti, P.C., Richter, D.J., Lavery, T., Kouyoumjian, R., Farhadian, S.F., Ward, R. *et al*. (2001). Linkage disequilibrium in the human genome. Nature *411*, 199–204.
8. Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M. *et al*. (2002). The structure of haplotype blocks in the human genome. Science *296*, 2225–2229.